Assessing the determinants of larval fish strike rates using computer vision

Shir Bar[a,b,*], Liraz Levy[b], Shai Avidan[c,**] and Roi Holzman[a,b,**]

[a] School of Zoology, The George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel
[b] The Interuniversity Institute for Marine Sciences, Eilat, Israel
[c] School of Electrical Engineering, The Iby and Aladar Fleischman Faculty of Engineering, Tel Aviv University, Tel Aviv, Israel

ARTICLE INFO

Keywords:
Machine vision
Action classification
Automated video analysis
Feeding behavior

ABSTRACT

Measuring behaviors that affect fitness is a critical task in the study of ecology and evolution. Behaviors such as feeding, fleeing predators, fighting conspecifics and mating are critical to an individuals' fitness but are often unpredictable in space and time and can be rare in natural and experimental systems. Sparsely occuring behaviors are therefore difficult to quantify, and extracting them from video streams is extremely time consuming. In this study, we use a case study of larval fish feeding, a sparse behavior that is critical to larval survival, to demonstrate how an AI-assisted system can be applied to overcome the problem of quantifying sparsely occuring events. We deployed our system in aquaculture rearing ponds to directly estimate the strike rate of larval fish outside the laboratory for the first time, and assess the effects of environmental factors on these rates. Our analysis pipeline far surpassed the performance of manual annotation, both in terms of time efficiency and its ability to retrieve feeding strikes. We found that strike rates were similar and low across age groups, irrespective of pH and oxygen levels. However, strike rates increased significantly with increasing temperature. Our system allowed probing into the biology of a sparsely occuring behavior with unprecedented efficiency. However, it revealed that analyzing rare behaviors requires further development of research methodologies suited for low sample sizes and highly imbalanced data.

1. Introduction

The concept of fitness is a hallmark in biology and is considered the driving force of evolution (Schluter, 2001). It is defined as the relative contribution to the gene pool of the next generation by an individual (Schluter, 2001). Despite its importance, fitness is a difficult metric to quantify directly. To measure fitness outside the laboratory, studies often use field enclosures (Martin and Wainwright, 2013) or large mesocosms (Arnegard et al., 2014; Bassar et al., 2015). In practice, fitness is often approximated by metrics such as reproductive capability, social dominance, or lifespan (McGraw and Caswell, 1996; Ariew and Lewontin, 2004; Van de Walle et al., 2022; Viblanc et al., 2022). Accordingly, mating, feeding, avoiding predators, and engaging in intra-specific interactions have a large effect on an animal's fitness and are thought to be under strong natural selection.

Disproportionately to their importance, quantifying these behaviors can be exceedingly difficult. This is because in many cases mating, hunting, avoiding predators, and engaging in intra-specific interactions can be rare (i.e., sparse). Here, we define "sparse behaviors" as behaviors that are (1) unpredictable in space and time, and (2) their natural frequency is low enough to make it difficult to detect and extract them from continuous observational data. Useful criteria to define "low frequency" can be the commonly utilized threshold of Fisher's P value (probability to randomly select the behavior of interest from the pool is <0.05) or the 3- and 6-sigma rules for outlier detection (Cook et al., 2021). However what exactly constitutes a "sparse behavior" will differ with the nature of the data, the way it is collected, the study system, and how distinctive the behaviors are (Cook et al., 2021).

*Corresponding author
**Equal contribution
shirbar@tauex.tau.ac.il (S. Bar); avidan@tau.eng.ac.il (S. Avidan); holzman@tau.ac.il (R. Holzman)
ORCID(s):

The advent of imaging technology has boosted the use of videos and still images for modern observation-based ecological and behavioral research (Tuia et al., 2022; Datta et al., 2019). These methods have the advantage of being relatively cheap, non-intrusive, and easily scalable. Unlike individual-mounted sensors, they do not require capturing or handling animals. However, raw data obtained by videography and automated photography is amassing quickly, while analysis capabilities remain a bottleneck; specifically for sparse behaviors. For example, Bessa et al. (2022) deployed 56 trail cameras for roughly 18 months to study the cultural behaviors of four chimpanzee communities. The resulting dataset contained $\sim$ 4000 video clips, but only $\sim$ 1150 contained chimpanzee behaviors. Of these, only 203 ($\sim$ 5%) of the clips featured behaviors that were of immediate interest to researchers. The mean number of video clips for these behavioral classes of interest was low, with a mean of 14 clips (range 1-64 clips). Such low ratios of gain-to-effort provide a strong incentive to automate the detection of sparse fitness-determining events in laboratory, field, and mesocosm studies (Hanscom et al., 2023).

Computer vision, the study of extracting information from imagery, can provide solutions to ecologists (Weinstein, 2018; Tuia et al., 2022; Christin et al., 2019). But, the small sample sizes that plague the study of fitness-related behaviors hinder not only biological research but also most modern computer vision methods. In particular, deep learning models (LeCun et al., 2015), which can provide the key to automation, are infamously "data hungry" (Christin et al., 2019; Aggarwal et al., 2018). To analyze data using a deep learning model, the model has to be trained first, using observations that represent the behaviors of interest, in a visually-similar environment. Typically, these training datasets consist of at least a few hundred (and more commonly 1000s) examples per "action class", or behavior (e.g., Carreira and Zisserman, 2017). This leads to a paradoxical situation whereby a researcher wanting to automate their research pipeline will likely collect sufficient data that would allow them to answer their biological questions before collecting enough data to train a deep learning model. Thus, it is challenging to apply novel computer vision solutions to biological data. This is particularly true for datasets of sparsely occurring behaviors that, as shown above (Bessa et al., 2022), are typically small and highly variable.

In this paper, we propose a deep learning pipeline for the detection of sparsely occurring behaviors in continuous video data (Fig. 1). We focus on one model behavior, that of larval fish feeding in mesocosms under variable environmental and filming conditions. Larval feeding is known to be rare (few prey items per hour per fish), especially during early ontogeny (China and Holzman, 2014; China et al., 2017; Shamur et al., 2016). Moreover, because the larvae and the prey are freely moving in a large volume of water, feeding attempts are unpredictable in space and time. Lastly, larval feeding is known to be critical to larval growth and survivorship, and hence can be considered a fitness-determining trait (China and Holzman, 2014).

Although we filmed within a set of mesocosms, we faced considerable challenges that are typical of uncontrolled conditions (Barbedo, 2022). These included fluctuating water turbidity and lighting, prey and larval fish density, currents, and turbulence. In addition, because we were filming a small volume within a much larger tank, we experienced individuals occluding one another, moving in and out of the frame, in and out of focus, and appearing in variable orientations with respect to the camera. These challenges are not unique to underwater mesocosms and are also reported in studies applying computer vision to study animal behavior and movement in aviaries (Xiao et al., 2023), zoos, and terrestrial enclosures (Joska et al., 2021; Labuguen et al., 2021). These features are usually avoided in the laboratory. We discuss our use case as an intermediate step from estimating sparse behaviors in the laboratory and in the field. We share some of the practical issues we were confronted with when using AI models to quantify a rare behavior under variable visual conditions and provide insights as to how to address them. In particular, we discuss how to evaluate the performance of binary classifiers on imbalanced data and highlight issues with common performance metrics. Lastly, we show how using our method we were able to achieve not only a speedup in time efficiency and

reduction of work effort but also retrieve twice the amount of behavioral events of interest when compared with manual analysis.

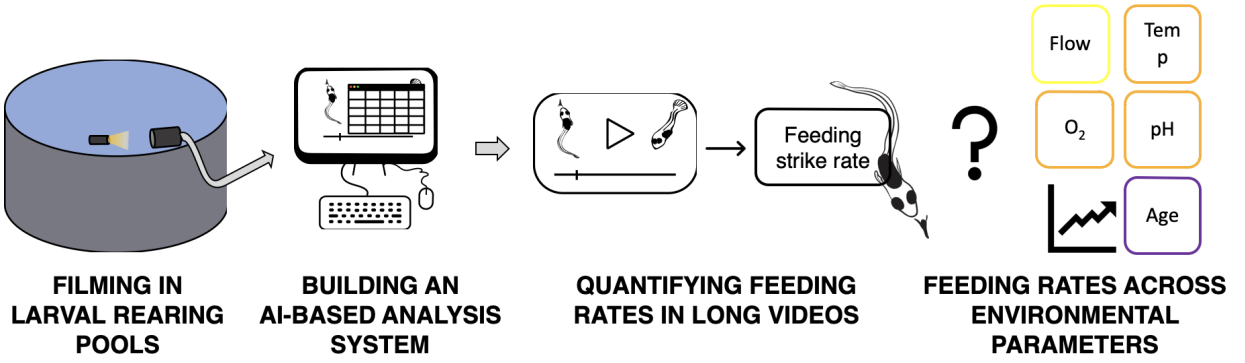

Figure 1: Overview of the study workflow. Deployed in an aquaculture rearing pool (left), our specialized system yielded high-speed videos of larval fish behavior. We manually reviewed these, searching for larval feeding strikes. We then developed an AI-assisted analysis pipeline and benchmarked it against the manual review method. To complete the pipeline, we show that we can reliably quantify feeding strike rates and display a use case for investigating the link between strike rates and environmental variables.

1.1. Examples of quantifying sparse behaviors span across taxa and study systems

Wrangham and Van Zinnicq Bergmann Riss (1990) quantified predation rates by two chimpanzee communities over the course of 3 years, and invested more than 14.5K man hours of 30 researchers in order to collect observations. The resulting predation rates were too low to estimate in female chimpanzees and were estimated for males as 0.31 kills per male per 100 hours with a sample size of 17 individuals. In cooperatively breeding manakins, extra-paired copulations were never observed in 3,400 hrs of observation and 19,221 hours of video recordings, despite genetic evidence for such cheating events (Boyle and Shogren, 2019). A study on beach-hunting, a rare feeding strategy in dolphins, invested over 10 years of observation to describe this behavior only in four maternal lineages (a total of 23 focal individuals) within the local populations (Sargeant et al., 2005). Other examples of rare behaviors and events include birth rate in Savannah Baboons (McLean et al., 2019) (N=199 from a 47-year-long database), feeding in snakes (Hanscom et al., 2023), proportional prey tracking in falcons (Cook et al., 2021), and hybridization in Galapagos finches (Cook et al., 2021). These low rates of occurrence incur a high cost for data collection and thus limit the investigation of new populations and study sites. This limitation makes it difficult to understand the effect of variation in environmental conditions on the measured behaviors and limits our ability to generalize beyond populations where these behaviors are easy to measure.

When the frequency of the behavior of interest is low (e.g. « 1%), extracting and curating a dataset suitable for statistical analysis becomes time-consuming and challenging. One common solution mainly used for biodiversity estimates in marine systems is Baited Remote Underwater Videos (BRUV). The baits in the systems are meant to lure fish in front of the camera setup. However, such systems are biased towards specific species, and since they elicit behaviors artificially, they are not suited to study the rates and distribution of naturally-occuring sparse behaviors (Sheehan et al., 2020). Another solution involves triggering the camera when a particular event of interest occurs, for example motion as in camera traps (e.g., Bessa et al., 2022), or using onboard object detection (e.g., Coro and Walsh, 2021). Such systems are useful when the organisms are rarely detected (e.g., unbaited camera in the deep sea Coro and Walsh, 2021) but not when a frequently encountered animal rarely performs a behavior of interest.

1.2. Larval fish feeding during early ontogeny

Most marine fish reproduce by external fertilization, releasing small (1 mm diameter) eggs into the water column. The embryos develop and hatch in the open ocean, and transition to exogenous feeding after a few days. This transition and the period following it are considered "the critical period" of larval fish, due to the extremely high mortality rates experienced by the larvae. It is estimated that >95% of the larvae die within a couple of weeks (Houde and Schekter, 1980; Hjort, 1914). Such mortality has far-reaching effects on the dynamics of natural populations and commercial fish stocks because small changes in survivorship rates translate to a large number of adults in the population. High rates of mortality are also a general phenomenon in the commercial rearing of marine fish larvae in aquaculture, where conditions are supposedly optimal. Mortality rates of »70% of the cohort are not uncommon (Shields, 2001) even when optimizing environmental factors such as temperature, oxygen, pH, and prey types. Despite decades of research on larval mortality during the critical period, there is no consensus regarding the relative importance of the processes that affect the magnitude of this mortality nor regarding the relative effects of environmental factors on it (Houde, 2008; Pepin, 2023). Yet, starvation is clearly a major mortality agent for larvae, and the larvae's ability to feed is a critical component that determines their individual survivorship.

Larval fish capture their prey using suction feeding. In this feeding mode, the fish abruptly open the mouth and expand the buccal cavity to generate an inflow of water into the mouth that carries the prey in with it (Holzman et al., 2015; Yaniv et al., 2014; China and Holzman, 2014). Suction-feeding is an extremely rapid behavior, with events typically taking less than 40 milliseconds from the start of mouth opening to mouth closing. Larval fish are minuscule, hatching at a body length of ∼3 mm, and growing to ∼10 mm at 30 days post-hatching (DPH). Correspondingly, both their mouth and prey diameter measure ca. 0.1-1 mm. The size of the animals, their speed, and erratic 3D motion, therefore, necessitated the development of a specialized filming system (Yúfera and Darias, 2007; Crespi and New, 2009; China and Holzman, 2014; Kamacı et al., 2005; Shamur et al., 2016).

Direct visualization of larval feeding under laboratory conditions brought about new insights into the mechanisms of feeding success (China and Holzman, 2014; China et al., 2017) and the mechanism of prey selectivity (Sommerfeld and Holzman, 2019). These studies revealed that the outcome of their interaction with prey is governed by the hydrodynamic regime in which they dwell and that the overall rate of successful prey capture was low, but increased with the age and size of the larvae. Quantification of hunger-related neuropeptides revealed that first feeding (8-10 DPH) larvae experience chronic hunger even under high prey densities, and experimental manipulation of the hydrodynamic regime in which the larvae dwell directly affects their level of starvation (as indicated by neuropeptide secretion; Koch et al., 2019).

However, direct observations on larval feeding are restricted to fish swimming in well-plates, petri-dishes, or small aquaria. The feeding rates observed under these conditions probably do not represent the rates typical of large-scale tanks and mesocosms due to prevailing wall effects, low spatial variability, and near-static water in the laboratory (Englund and Cooper, 2003; MacKenzie et al., 1990). To the best of our knowledge, there are no direct estimates of feeding rates of larval fish outside the laboratory, despite their potential importance for understanding larval condition and survivorship. Studies conducted in large mesocosms (e.g., larval rearing pools) and in the open ocean focused on indirect measures of feeding performance such as gut content and stable isotope analysis (Pepin, 2023; Schlechtriem et al., 2004). These methods provide an estimate of nutritional state rather than a mechanistic understanding of the determinants of feeding rates, as they cannot suggest a cause for the starvation nor inform researchers of the number of attempts (and effort) that preceded successful captures. Additionally, gut content is problematic because evacuation times are highly variable and contingent upon a multitude of intrinsic and extrinsic variables. Stable isotope analysis,

measurements of RNA/DNA ratios, and gene expression studies all provide useful information but suffer similar limitations (Tanaka et al., 2008; Foley et al., 2016; Buckley, 1984; Yúfera et al., 2018; Schlechtriem et al., 2004).

2. Methods

In this section, we present the development of an analysis pipeline for the detection of larval fish behavior in aquaculture rearing pools, from the initial infrastructure to the deep learning models (Fig. 1). We first briefly describe the study system and our novel imaging system designed to visualize millimeter-sized, fast-moving organisms underwater (section 2.1). We explain the AI-assisted analysis pipeline for the detection of the feeding strike behavior of larval fish and its evaluation (section 2.2). Here, we also expand on the challenges of training using very small sample sizes, compared with regular computer vision datasets, and how we chose to tackle them (section 2.2.1). Finally, we expand on our statistical analysis, to ascertain our measurements (section 2.3). An in-depth description of the training of the main powerhouse of our analysis pipeline, namely the classification module, is provided in the supplementary material (section S3) and our code.

2.1. Study system, data acquisition, and annotation

The feeding behavior of larval sea-bream *Sparus aurata* was recorded for 17 cohorts over the course of 18 months in eight rearing tanks located in a commercial aquaculture facility (ARDAG hatchery, Eilat, Israel). We deployed a submerged camera system in the tanks to record the activity of freely-behaving larval fish, from an age of 8 DPH to 30 DPH. In this facility, larvae are reared in large circular tanks (4 m in diameter and ∼1.5 m deep, Fig. 2a) under controlled temperature and oxygen conditions. Cohorts of $\sim 10^6$ eggs are introduced into each tank, where they hatch and are grown until they metamorphose to adult-like morphology at ∼35 DPH. Larvae are fed twice a day with various food types (Rotifers, Artemia, and pellets) according to their age and size. Note that, here, controlled does not mean constant. Rearing protocols typically indicate a range of acceptable values to be maintained for each parameter (e.g., 80-100 % for O_2), and changes in temperature, oxygen, and prey type may appear at different stages of the ontogeny. Additionally, as a large-scale commercial facility, conditions are not as strictly maintained as in the laboratory, and deviations from desired conditions are common.

Filming was done using a high-speed monochrome camera (Optronics CP70-2-M/C-1000) enclosed in an underwater housing, submerged to ∼0.2-0.3 m depth, and connected to a computer equipped with a frame-grabber (Cyton Quad Channel CoaXPress frame grabber CXP6) and a GPU (NVIDIA GeForce RTX)(Fig. 2 a,b). The system recorded long high-speed videos (20-30 min) at 500 or 750 (N=191 and N=32, respectively) frames per second (fps) and a resolution of 1920×1080 pixels. During the course of the study, the lower frame rate was found to be sufficient for identifying the feeding attempts, while being more reliable in terms of video quality.

The camera was equipped with a Navitar 6000 ultra-zoom lens providing 2:1 magnification (i.e., 2 mm in the real world is mapped on 1 mm of the camera's sensor) with a large depth-of-field of 50 mm. A battery-operated SCUBA flashlight (Scubatec US15 LED 10 Watt) provided backlight illumination essential for filming in high frame rates and under such optical constraints. See supplementary material section S1.1 "illumination system" for a discussion of possible effects on larval behavior.

The setup enabled visualizing events within a volume of $40 \times 60 \times 50$ mm ($H \times W \times D$). At the beginning of each filming day, we recorded the temperature and oxygen concentration in the pool using a HOBO oxygen meter (HOBO U26-001 Dissolved Oxygen, Onset Computer Corp). The pH of the water was measured using a pH electrode.

Overall, we obtained and reviewed over 77.73 hours of high-speed videos in 223 videos; accumulating to ∼ 146.25 million frames. To manually annotate the sparse feeding events, a trained observer watched each video at 10-15 FPS

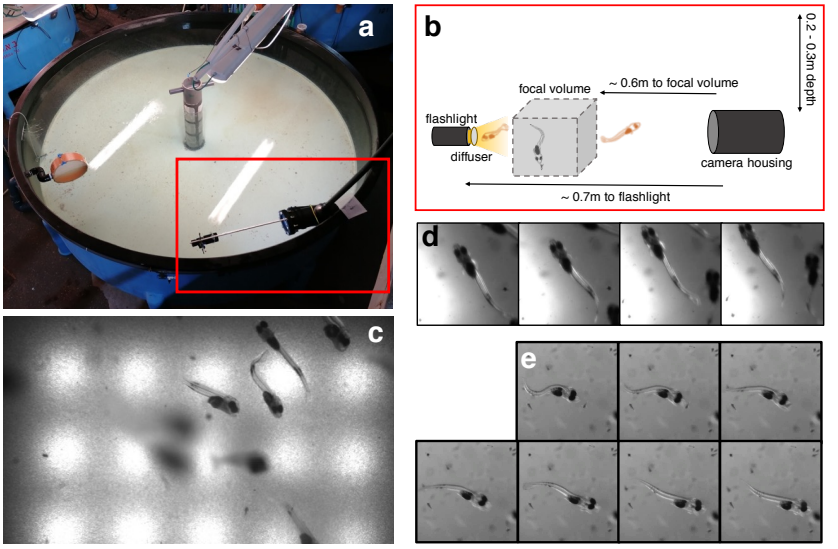

Figure 2: The camera system and data acquisition setup. (a) the camera system in a fish larvae rearing pool at the ARDAG hatchery, Eilat; (b) a schematic depiction of the submerged camera setup, including a representation of the focal volume, where fish appear sharpest; note that this volume is a mere 0.002 % of the pool (c) an example of a full video frame acquired using the system. Larvae appear in various levels of sharpness within the frame, see Fig. S1 in supplementary material for more examples; (d) an example of a "swim" sequence from the dataset (see clip Video S1); and (e) a representative sequence of a "strike" event from the dataset (see clip Video S2).

and noted the time and coordinates of all feeding strikes. Strikes were defined as events that combined a rapid forward lunge and opening of the mouth (Fig. 2e). Such strikes are visually distinct and represent high-effort prey-acquisition attempts that are likely to be successful (China et al., 2017). We manually annotated 149 full-length videos (mean duration $= 20.92 \pm 12.93$ min). We estimate that annotating this dataset took a minimum of 2,235 work hours; or about 50 min work per minute of video. This is an underestimation as the observers often had to re-play the video to ascertain an observation.

2.2. Analysis pipeline

We analyzed our videos using a computer vision pipeline, consisting of two deep learning models: a fish detector (Faster-R-CNN, Girshick, 2015) and an action classifier. We trained and evaluated each model separately (sections S3 and S4 in the supplementary material). After training, and prior to the application of this pipeline to long, uncut videos, we evaluated the performance of several action classifiers on a test set in order to select the best-performing model for the pipeline. Below, we discuss each of the steps above. We focus on consideration for training using a small sample size, classifier evaluation, and the application of the whole pipeline to our untrimmed video data. See supplementary material (section S3) for more technical details on classifiers' training and their performance

2.2.1. Training binary classifiers under a low data regime

Our training data contained 71 non-strikes (swims; negative class) and 66 strikes (positive class) and was limited by the number of strike samples available to us at the time. This is a modest number for training, posing a challenge with training the classifiers. However, this is a common difficulty in ecological datasets, especially when the behaviors of interest are scarce. We tackled this data scarcity using complementary approaches: A highly curated and balanced training set, transfer learning (Tan et al., 2018), intensive data augmentations tailored for our dataset, and integration of our prior biological knowledge as additional input to the model.

We chose to train on a balanced dataset, keeping the number of non-strikes samples similar to that of the strikes. Other methods for training on imbalanced data include class resampling (undersampling the majority class, or upsampling the minority class), weighted loss, or focal loss, but were not used in this study. To make the most out of the data, we used only samples containing high-quality clips of single fish performing a single action, with each clip being tightly cropped around the fish. The duration of each strike clip was cut temporally 10 frames before the fish's mouth opened and 5 frames after the mouth closed. We contrasted the 'strike' class with only clips of fish performing routine undulatory swims. This was also done in order to maximize the model's ability to learn the abrupt kinematics of strike events. Details on the training set and its partitions are available in the supplementary material (Table S1).

We used transfer learning to leverage open-source models trained on large image datasets. Essentially, instead of initializing model parameters randomly, we start from the trained model and then fine-tune it for our classification task. Please see our code for further details.

To assess the contribution of pre-training to model performance, we trained a SlowFast network from scratch on our dataset. We also tested whether using a model pre-trained on a different dataset might improve performance. For this, we compared a couple of SlowFast networks pre-trained on different datasets. The first network was pre-trained on the Kinetics-400 dataset (Carreira and Zisserman, 2017) - a dataset of 400 human action classes, with over 400 clips per class. A second network was pre-trained on the Something-SomethingV2 dataset (SSv2) (Goyal et al., 2017), after being trained on Kinetics. We were inspired by works such as Mathis et al. (2018) which showed that pre-training on human pose data is beneficial when learning animal pose, in spite of the difference in domains. We chose the SSv2 dataset because recent work suggested it encourages the learning of more dynamic, temporal-related features (Kowal et al., 2022). In all experiments, we used the publicly available models and weights in the PyTorchVideo repository (Fan et al., 2021; Paszke et al., 2019). We fine-tune all weights in the model for 50 epochs, full details are provided in (Bar et al., Unpublished results) and our code.

Our training dataset, though small, showed a diversity of visual conditions; main differences in lighting intensity, and degree of blurriness of the fish (Fig. S1). To encourage model generalization over these conditions and to enhance the number of samples in our dataset, we randomly applied augmentations (modifications) to the intensity values of clips, varying the degree of brightness, and augmented the sharpness of clips by randomly applying Gaussian blur to samples during training.

We integrated our knowledge of the biology and behavior of the fish to generate an additional channel of information in our clips. We exploited the fact that "strike" behavior is characterized by abrupt movements, while "swim" behavior is typically a smooth undulatory movement. These differences are expected to affect the rate at which pixels change their intensity values throughout the clip, with "strike" pixels showing areas of higher variance. To capture this, we calculated a single variance image of the entire clip (see Fig. S2) and used it as additional input to the model. The variance image was duplicated along the temporal axis and stored as a third channel, alongside two duplicate channels of the clip's monochrome sequence. We note that given a larger dataset, we would expect our classification module to learn such features independently, making this additional channel superfluous. However, in light of our low data regime, we considered that this manipulation would inform the classifier and enhance learning.

2.2.2. Classifier evaluation

After successfully training and comparing several different classifiers, we evaluated the two best-performing models, both variants of SlowFast action detection network (Feichtenhofer et al., 2019), on an imbalanced test set. This test set was created to emulate the extreme class imbalance and challenging visual conditions (e.g., overexposure, occlusions, blurred images) prevalent in untrimmed video sequences.

The test set consists of short clips and has two behavior classes; non-strike (n=4,500) and striking on prey (n=63). The non-strike class featured mostly routine swimming behavior, but also turns, acceleration and deceleration, and some unusual abrupt movements that we associated with spitting prey. We treat the strike class as the positive and the non-strike as the negative class.

The action classifiers assign a "strike score" to each clip, indicating the probability of the clip belonging to the positive class. To assign clips to action classes, one needs to select a threshold score above which clips are considered strikes and under which clips are considered non-strikes. Ideally, this "decision threshold" would minimize the number of false predictions by the classifier. Such false predictions can be divided into two cases: (1) false positives, i.e., non-strike being classified as strikes; and (2) false negatives, i.e., strikes being classified as non-strike.

As customary in binary classification evaluation, we use the Receiver Operator Curve (ROC) (Hanley et al., 1989), to obtain an estimate of the overall classifier performance across the entire range of decision thresholds. The area under this curve (AuROC) is often used to give a single score to the quality of the classifier. We also evaluated the classifier using the Precision-Recall Curve (PRC), following best practice for imbalance data (Saito and Rehmsmeier, 2015), and used the area under this curve (AuPRC) as an additional quality score. PRC is sensitive to the class imbalance of the dataset, with expected curves changing according to the positive class percent in the data (Brabec et al., 2020), and therefore caution should be taken when extrapolating the performance for datasets of different class imbalances. All metrics were calculated using the precrec package in R (Saito and Rehmsmeier, 2017).

2.2.3. Pipeline application

The best-performing classifier on the test set, the SlowFast network pre-trained on the SSv2 dataset, was used in the pipeline in conjunction with our trained Faster-R-CNN fish detector (supplementary material section S4). The pipeline was applied to the first five minutes of each of our 223 videos. We applied it in a sliding window fashion such that a focal frame is sampled every 60 frames throughout each 5-minute sequence. For each focal frame, the detector was applied to locate all the fish in the frame (see Fig. 3). Around each of these detections, we created a short cropped clip. The clips were centered around each detection, with the cropping size determined according to the typical size of the fish in each video (range: 250-650 pixels). Temporally, each clip consisted of a $\pm$40 frame window of the cropped area (total of 80 frames). For example, a larva detected in a focal frame (e.g., frame 100) was imaged for 80 frames, extending from frame 60 to frame 140. If the larva was also detected in the next focal frame (e.g., frame 160), it would be imaged again from frames 120 to 200. This sliding window sampling scheme creates a temporal overlap of 20 frames between clips from subsequent focal frames. This overlap was intended to reduce edge effects (e.g., strikes that occurred at the very beginning or end of a clip). Each resulting clip was then fed to the action classifier. the duration of the extracted clips was $\sim$ double the duration of a feeding event ($\sim$ 46 frames on average), ensuring that we do not miss events due to sampling.

Since false positive rates were found to be high on the test set, we manually reviewed all clips that were predicted to be strikes by our pipeline (i.e., were above the classifier's decision threshold) to ascertain the predictions. This is why we consider the system an AI-assisted system rather than a fully automatic AI pipeline.

2.3. Statistical analysis

To determine whether our sampling of a small fraction of the population was sufficient to estimate the average strike rates in the population, we performed a bootstrap analysis on the coefficients of a zero-inflated Negative Binomial model (Zeileis et al., 2008). This model discerns between two types of zeros: (1) sampling zeros, which represent the portion of larvae that did not feed but were visible in the frame, and (2) structural zeros, which represent cases where feeding may have occurred, but was not visualized and hence not counted. We modeled the feeding strikes as a

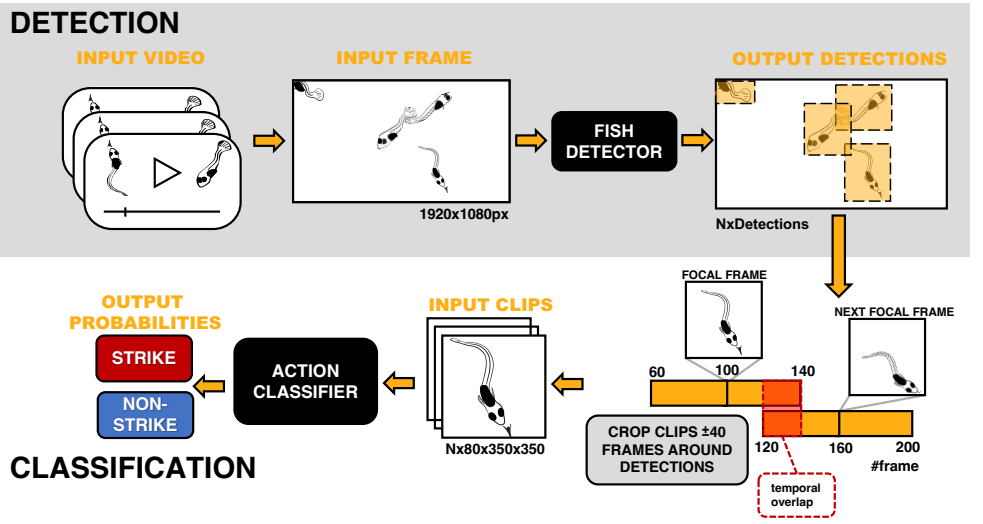

Figure 3: A diagram of the analysis pipeline. The flow depicted in this diagram, starting in the upper left corner, shows how we applied our analysis pipeline to long video sequences. Using a sliding window approach we sample a focal frame every 60 frames and generate fish detections on that particular frame. Using these detections, we crop short clips centered around individual fish (±40 frames, temporally overlapping between adjacent focal frames) and feed these into an action classifier trained to predict whether the fish is performing a feeding strike. The pipeline reduces the problem of reviewing long sequences with large frames containing multitudes of fish, to that of reviewing short clips mostly centered around a single fish.

function of age, with the zero-inflated portion of the model depending on the mean density of fish per frame. This is because we reasoned that the main process governing the probability of strike visualization is the density of fish visible in the frame. To calculate the mean density of fish, we sampled 20 random frames uniformly from the length of the video and averaged the number of detections by our fish detector across frames. We chose to bootstrap the coefficients of the model rather than the raw feeding strike rate to account for the zero-inflated nature of the data. If we were to directly bootstrap the rates we would have been bootstrapping the stochastic processes (noisy) governing fish density and not our behavior of interest. Details on the bootstrap procedure and the zero-inflated model used can be found in the supplementary material (section S5).

We next asked whether larval age, water temperature, oxygen concentration, and pH (independent variables) affect the observed strike rates (dependent variables). Because many of our videos featured no feeding events, we used a zero-inflated model (as explained above) to test this hypothesis. For this analysis, we use all the strike events identified during classifier evaluation, i.e., those identified by the final AI-assisted pipeline, as well as ones that were identified by the annotators during manual analysis and by classifiers with inferior performance (a total of n=128 strike events). Similar results were obtained using only strikes detected by the pipeline (see supplementary section S5.4) Four videos for which we had no temperature data were omitted from this analysis. All statistical analyses were implemented in R; zero-inflated models were computed using the *pscl package* (Zeileis et al., 2008).

3. Results

3.1. Classifier performance on the test set

We used our test set of 4,563 clips to assess the performance of different classifiers and to determine a decision threshold to assign clips to action classes. The ability of our best-performing classifier to detect true strikes was very high, as indicated by the AuROC of 0.97 ± 0.003. However, all classifiers had intermediate precision, meaning that

most of the clips assigned to the "strike" action class featured non-strike behaviors (were false positives). Accordingly, the AuPRC of our classifier was 0.66 ± 0.015. See supplementary material section S3.4 for full details on the classifier evaluation. This intermediate precision is a corollary of the extreme rarity of the positive (strike) class. When selecting a decision threshold for data evaluation (see 2.2.2), we set the threshold to 0.75 (see supplementary material section S3.6). Based on the distributions of strike scores in our test dataset, this threshold was expected to enable a recall of 80% of the strike events.

3.2. Applying the AI-assisted pipeline to untrimmed videos

We applied our AI-assisted pipeline to 149 high-speed videos, which had also been manually annotated. The pipeline generated 414,595 clips, of which 5.33% were above the decision threshold. We manually reviewed these upper-percentile clips and annotated them according to their action class. For each video, we reviewed either the first 500 clips or the entire 5 minutes, whichever came first. This AI-assisted pipeline outperformed the trained observers in its ability to detect feeding events (Fig. 4) and in terms of time efficiency. Our trained observers detected 27 strikes during the first 5 minutes of these videos, whereas the AI-assisted pipeline retrieved 47 strikes. Of these, 20 were strikes detected by the trained observers. Thus, the AI pipeline missed 7 events detected by the observers, whereas the observers missed 27 events detected by the pipeline. If we consider the pool of all events detected by both methods in the 149 5-min long videos (N=54), we can compare the recall of the human observers (50%) to that of the AI-assisted pipeline (87%; Fig. 4). The difference in performance between the two methods was found to be statistically significant (Pearson's Chi-Square; $\chi^2 = 23.5$, $df = 2$, $p << 0.001$).

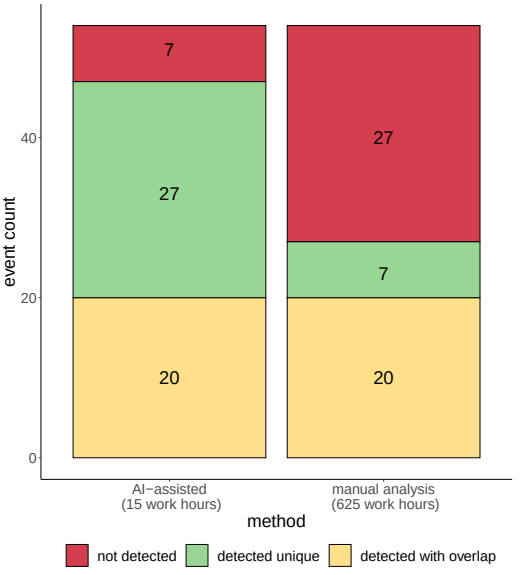

Figure 4: Pipeline performance on long videos. A comparison of the strike events detected and missed by our AI-assisted pipeline (left) and the manual analysis (right) if we consider the pool of all events detected by both methods in 149 videos (N=54). An estimate of work hours is specified in parenthesis. Events could either be detected by both methods (yellow), detected (green) or missed (red) uniquely by the focal method. Numbers within each block are the number of events. The sum of the green and yellow areas in each bar is the total events detected by that focal method. The AI-assisted pipeline detected 42% more events compared with the manual analysis.

In terms of time efficiency, the pipeline was designed to run offline, as such execution times are roughly one hour per video ($average \pm SD$ of 71 ± 79 min per video, median 49.5 minutes). Fully manual annotation of the 149 5-min long videos took roughly $\sim$ 625 work hours whereas reviewing the upper-percentile clips as mentioned above took $\sim$ 15 observer work hours.

The performance advantage of the AI-assisted pipeline allowed us to analyze 74 additional videos that were not manually annotated due to time and budget constraints. In these videos, the pipeline generated 197,717 clips, of which 9.54% were above the decision threshold. A manual review of these upper-percentile clips identified 25 additional strikes.

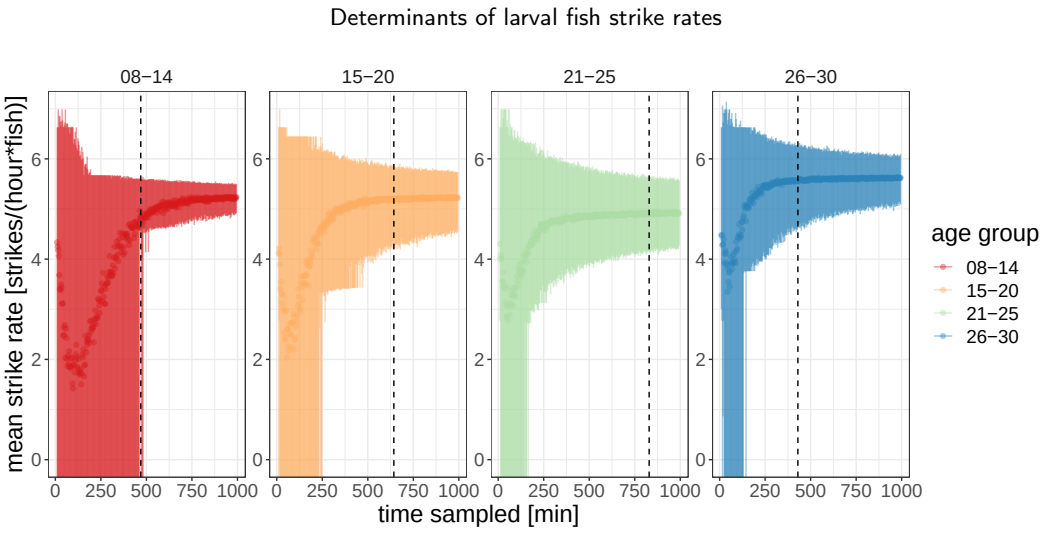

Figure 5: Accuracy of feeding strike rate estimates. Presented are the bootstrapped value for a Zero-inflated Negative Binomial model describing the relationship between strike rate and age while accounting for zero inflation caused by low fish densities in the imaged volume. Each panel shows the strike rate estimated for a given age group. The x-axis represented the sample size (in minutes of video) for which the model was calculated. The round markers are the mean of 1000 bootstrapped samples. The colored lines are the 95 % confidence intervals (CI). The dashed black line represents the point where the CIs are 15 % of the value of the mean. All groups converge to a stable estimate well before our maximal sample size; suggesting we can accurately measure strike rates for our population.

3.3. Statistical analysis - are we accurately estimating strike rates?

The bootstrap analysis revealed that using our pipeline with the best classifier, we sampled sufficiently to estimate strike rates for all age groups with a confidence interval of $< 15\%$ of the mean (Fig. 5). Another noticeable effect is the noisier convergence of the two younger age groups (8-14,15-20), which also had higher levels of zero inflation. All in all these results suggest that, by using appropriate statistical models, we can accurately estimate the strike rates of the fish in our study system using our analysis pipeline.

3.4. Effect of environmental parameters

The mean estimated strike rate was 8.1 ± 3.45 prey items per hour per fish. Strike rates doubled as water temperature increased from 19 to 22 degrees (zero-inflated model, $p<0.005$), but were unaffected by larval age, oxygen concentration, and pH (zero-inflated negative binomial model, $p>0.05$, supplementary Table S5; Fig. 6).

4. Discussion

In this study, we present an AI-assisted method for detecting rare behaviors of freely behaving fish larvae (Fig. 3). We successfully trained our deep learning models using limited data (Fig. S4, Fig. 4), to produce an analysis pipeline that far surpasses the performance of manual annotation by experts, not only in terms of time efficiency but also in its ability to retrieve events of interest (Fig. 4). We used our data to directly estimate the strike rates of larval fish in large mesocosms outside the laboratory (Fig. 5) and to assess the effects of environmental factors, as maintened by an aquaculture rearing facility, on these rates (Fig. 6). Our analysis revealed that strike rates were similar and low across age groups, irrespective of pH and oxygen levels (Table S5). Strike rates doubled with a 3 degrees increase in water temperature (Fig. 6). Similar effects were reported in larval fish based on stomach content analysis (Cunha and Planas, 1995).

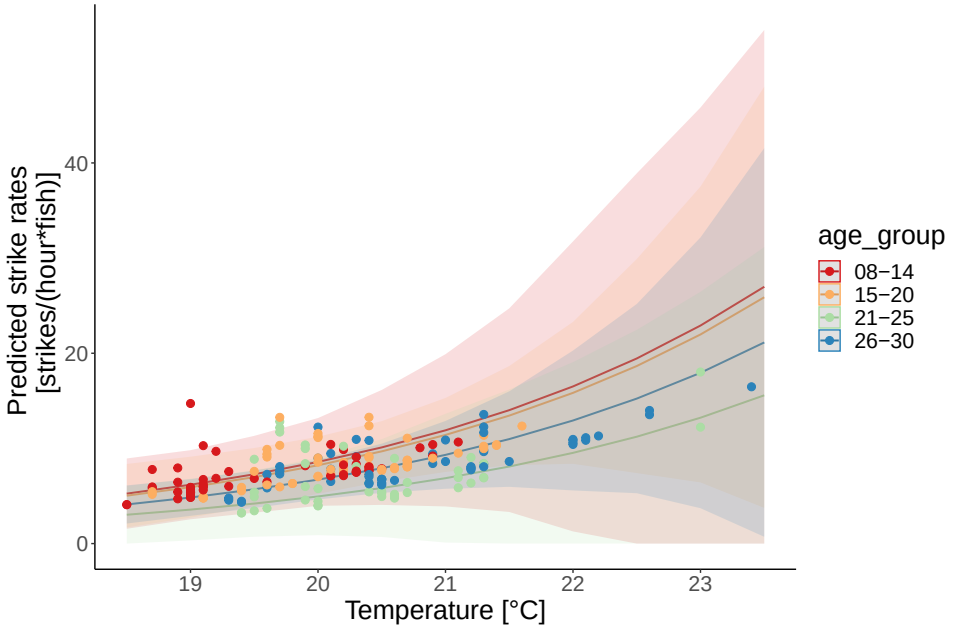

Figure 6: Modelled effect of temperature on strike rates. The strike rates (y-axis) predicted by a zero-inflated negative binomial model (round markers) for the observed temperature gradient (x-axis) are plotted for each age group (color). The predictions were calculated for the median values of oxygen, pH, and fish density. The curves are the model's fit across the temperature gradient. Shaded areas around each line are the 95 % confidence intervals. Temperature has a significant positive effect on strike rates, and differences between age groups were not significant. See Table S5 in the supplementary material.

Our pipeline, although not fully automatic yet, outperformed the unassisted manual analysis of untrimmed videos by expert human annotators (Fig. 4). Even when adding the computer processing time at the time budjet of the AI-assisted pipeline, we obtain a 77.7 % improvement in work hours (139 hours for the pipeline vs 625 hours for the annotators). By making an informed choice regarding its decision thresholds, we were able to reduce the work needed for manual review within the pipeline by $\sim$ 93.3%. Though selecting a high decision threshold (0.75) incurred a cost in terms of recall (only 74% othe events detected by observers were detected), we found that manual analysis resulted in even poorer recall. Since using our assisted scheme we found events missed by the observers, in fact the human annotators recall on the videos was only 50%, making our method not only faster but also better at event retrieval.

Evidence of reduced accuracy of human annotation in data acquired underwater in uncontrolled conditions was mainly attributed to low image quality stemming from the difficult filming environment (Barbedo, 2022). Additionally, this can be partially explained by the limited capacity of human cognition. Processing long, high-resolution frames containing multiple behaving fish is challenging and requires a meticulous review of each visible fish. A naive solution would be to film videos with a lower density of individuals, but as we are filming freely behaving animals as grown by the rearing facility, we could not modulate or control the density of fish imaged.

To the best of our knowledge, this is the first time that larval feeding strike rates were directly measured outside of the laboratory in large mesocosms. The rates we estimated are on the same order of magnitude as the feeding success rates documented in the laboratory by China and Holzman (2014). But, feeding success rates represent only some of the strikes, because 38.5-72.4% of attempts (depending on age, 9 - 23 DPH) constitute a failure to capture the prey (China et al., 2017; China and Holzman, 2014). Taking this into account, the strike rates we observed are substantially

lower than those observed in the laboratory; implying that the feeding rates in the rearing pools will also be markedly lower. However, our data do not reveal what factors limit the larvae's strike rate. Our statistical analysis indicates that oxygen concentration and pH levels did not significantly affect strike rates in the pools. Furthermore, pH and oxygen levels in the pools and the laboratory were similar, and within the recommended range for the species. The food concentrations in the laboratory were similar to that of the rearing pools, and in both locations seemed saturated. A major difference between the rearing pools and the laboratory was the occurrence of currents and turbulence, stemming from the vigorous aeration of the rearing pools. Strong turbulence is known to impede larval feeding and attempt rates (MacKenzie and Kiørboe, 2000) Unfortunately, we did not quantify the current and turbulence within the pools, therefore we cannot assess this hypothesis quantitatively.

4.1. Future directions and transferability to other systems

Ideally, a well-trained classifier should allow a fully-automated classification pipeline. However, we suggest that the limitations of our pipeline are an inherent problem stemming from the rarity of the events of interest. When sampling rare behavioral events from a background of more frequent ones, the resulting datasets are highly imbalanced, with the minor (positive) class being the behavioral event of interest (e.g., Bessa et al., 2022; Thomson et al., 2015). When applying our pipeline to full videos, the background class was 99.82% of the data reviewed. With strikes making up only 0.18% of the data, even a classifier with an error rate of 1.5% (erroneously assigning "non-strikes" to the "strikes" class) will assign similar numbers of "strikes" and "non-strikes" to the "strike" class and have low precision.

An emerging solution to the automatic detection of rare events might be the application of a different computer vision approach, namely an unsupervised anomaly detection algorithm. This method involves training an algorithm on the imagery of ”normal” events; defined as the class of events that are common but of no interest to the researcher. By definition, videos of freely-behaving animals are unlikely to contain rare events, therefore it is easy to generate a training set of routine events, with minimal annotation effort. An anomaly detection algorithm learns the distribution that defines the ”normal” events during training. At inference, it compares new imagery to this distribution space. Events that fall outside the learned space are considered ”anomalies”; these putative anomalies represent rare events that can be of biological importance. Applications of anomaly detection algorithms to animal behaviors are few and mostly based on low-dimensional sensor data (e.g., GPS, ECG, and motion sensors, Kiersztyn et al., 2022; Lenning et al., 2017; Haladjian et al., 2017; Cai et al., 2019). However, our study provides a dataset that could be used for training and evaluation of future anomaly detection pipelines on videos, a rich, high-dimensional data signal. We feel this direction, once properly tried and tested, can provide a tool for quantifying rare fitness-determining behaviors, hopefully alleviating data analysis bottlenecks.

Most computer vision systems require re-evaluation and training when transferring to a new environment. Specifically for our pipeline, it generalized well over a wide range of visual conditions. This is a good indication for its potential transferability, given proper training data. However, further adaptations are needed before applying our pipeline in the field, particularly given the yet high false positive amounts, which might render the search for scarce behaviors unfeasible. To use our system in a new environment one will likely need to train a detector and an action classifier with their own data. The amount of labeled data required will vary depending on the variability in visual conditions of the study sites, and how discernable are the behaviors of interest. Our larval fish detector can be replaced with any other off-the-shelf animal detector (e.g., MegaDetector in a camera trap setup Beery et al., Unpublished results) or a detector trained specifically for the dataset of interest. There are several popular code-bases making fine-tuning detectors accessible to people with little technical background (e.g., Wu et al., 2019). Detectors can be fine-tuned with a few hundred images, depending on the variability of visual conditions. Action classifiers require more attention, as the loading and feeding of video clips into models are not as straightforward. The training

of these classifiers is also more computationally intensive, and requires access to a computer with a GPU (we used a single NVIDIA R2090 with 12GB memory). For the classifiers, we also used open-source codebases (Fan et al., 2021; Feichtenhofer et al., 2019), all of our codes are available in accompanying GitHub repository. We showed that by selecting augmentations relevant to the data and adding a strong signal, a sort of prior, that focused the model on the motion of the input (the variance image) we were able to achieve a good recall of strike events. Similar use of such priors was also done by Kay et al. (2022), where they used the difference image between adjacent frames as additional input to the model. Selecting such data prior based on our knowledge of the behavior and its kinematics can thus greatly improve model performance, even on a low data budget.

Acknowledgments We thank Keren Perry and Tamara Gurevich for their indispensable help with acquiring, annotating and analyzing larval feeding behavior manually. We thank the Ardag Hatchery for the use of their facilities and continued support in our research. We also thank the entire Holzman lab for their help with the data labeling of the naturalistic dataset. We are indebted to Amir Jevnisek, Tal Perevolotsky and Anael Engel for inspiring brain-storming sessions. A special thanks to Moti Ohevia, Erez Levin, and the technical team of the Interuniversity Institute for Marine Science of Eilat for their help in the construction of the camera setups.

Statements and declarations The authors report no competing interests. This study was funded by the Israeli Ministry of Agriculture [grant number 13-37-0009], the Israeli Science Foundation [grant number ISF 592/22], Microsoft AI for Earth, the IDSI grant for cloud computing resources for graduate students, and The Colton Family Next Generation Technological Institute and the Miles Nadal institute for Technological Entrepreneurship at Tel Aviv University.

All datasets described in this paper will become available upon acceptance through an appropriate data repository. Full code is available in the paper's GitHub repository, as well as trained model weights. The authors state that they comply with the ethical standards as defined by the journal.

Author's contribution S.B: data curation, formal analysis, investigation, methodology, validation, visualization, writing—original draft and writing—review and editing; L.L: data acquisition, data curation, data analysis, methodology, writing—review and editing; S.A.: conceptualization, funding acquisition, methodology, resources, supervision and writing—original draft and writing—review and editing; R.H.: conceptualization, data curation, funding acquisition, methodology, resources, supervision, writing—original draft and writing—review and editing .

References

C. C. Aggarwal et al. *Neural networks and deep learning*. Springer Cham, 1 edition, 2018. URL https://doi.org/10.1007/978-3-319-94463-0.

A. Ariew and R. C. Lewontin. The confusions of fitness. *British Journal for the Philosophy of Science*, 55(2), 2004. URL https://doi.org/10.1093/bjps/55.2.347.

M. E. Arnegard, M. D. McGee, B. Matthews, K. B. Marchinko, G. L. Conte, S. Kabir, N. Bedford, S. Bergek, Y. F. Chan, F. C. Jones, et al. Genetics of ecological divergence during speciation. *Nature*, 511(7509):307–311, 2014. URL https://doi.org/10.1038/nature13301.

S. Bar, L. Levy, S. Avidan, and R. Holzman. Analysis of larval fish feeding behavior under naturalistic conditions. *bioRxiv*, Unpublished results. doi: https://doi.org/10.1101/2022.11.14.516417.

J. G. A. Barbedo. A review on the use of computer vision and artificial intelligence for fish recognition, monitoring, and management. *Fishes*, 7(6): 335, 2022. URL https://doi.org/10.3390/fishes7060335.

R. D. Bassar, T. Heatherly, M. C. Marshall, S. A. Thomas, A. S. Flecker, and D. N. Reznick. Population size-structure-dependent fitness and ecosystem consequences in trinidadian guppies. *Journal of Animal Ecology*, 84(4):955–968, 2015. URL https://doi.org/10.1111/1365-2656.12353.

S. Beery, D. Morris, and S. Yang. Efficient pipeline for camera trap image review. *arXiv preprint arXiv:1907.06772*, Unpublished results. URL https://doi.org/10.48550/arXiv.1907.06772.

J. Bessa, D. Biro, and K. Hockings. Inter-community behavioural variation confirmed through indirect methods in four neighbouring chimpanzee communities in cantanhez np, guinea-bissau. *Royal Society Open Science*, 9(2):211518, 2022. URL https://doi.org/10.1098/rsos.211518.

W. A. Boyle and E. H. Shogren. Sex and deception: a rare case of cheating in a lekking tropical bird. *Journal of Ethology*, 37(2):151–155, 2019. URL https://doi.org/10.1007/s10164-019-00592-8.

J. Brabec, T. Komárek, V. Franc, and L. Machlica. On model evaluation under non-constant class imbalance. In *International Conference on Computational Science*, pages 74–87. Springer, 2020. URL https://doi.org/10.1007/978-3-030-50423-6_6.

L. Buckley. Rna-dna ratio: an index of larval fish growth in the sea. *Marine Biology*, 80(3):291–298, 1984. URL https://doi.org/10.1007/BF00392824.

Y. Cai, L. Ma, and G. Liu. A night-time anomaly detection system of hog activities based on passive infrared detector. *Applied Engineering in Agriculture*, 35(4):481–493, 2019. URL https://doi.org/10.13031/aea.13007.

J. Carreira and A. Zisserman. Quo vadis, action recognition? a new model and the kinetics dataset. In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6299–6308, 2017. URL https://doi.org/10.1109/CVPR.2017.502.

V. China and R. Holzman. Hydrodynamic starvation in first-feeding larval fishes. *Proceedings of the National Academy of Sciences*, 111(22):8083–8088, 2014. doi: 10.1073/pnas.1323205111. URL https://doi.org/10.1073/pnas.1323205111.

V. China, L. Levy, A. Liberzon, T. Elmaliach, and R. Holzman. Hydrodynamic regime determines the feeding success of larval fish through the modulation of strike kinematics. *Proceedings of the Royal Society B: Biological Sciences*, 284(1853):20170235, 2017. doi: 10.1098/rspb.2017.0235. URL https://doi.org/10.1098/rspb.2017.0235.

S. Christin, É. Hervet, and N. Lecomte. Applications for deep learning in ecology. *Methods in Ecology and Evolution*, 10(10):1632–1644, 2019. doi: 10.1111/2041-210X.13256. URL https://doi.org/10.1111/2041-210X.13256.

C. N. Cook, A. R. Freeman, J. C. Liao, and L. A. Mangiamele. The philosophy of outliers: Reintegrating rare events into biological science. *Integrative and Comparative Biology*, 61(6):2191–2198, 2021. URL https://doi.org/10.1093/icb/icab166.

G. Coro and M. B. Walsh. An intelligent and cost-effective remote underwater video device for fish size monitoring. *Ecological Informatics*, 63:101311, 2021.

V. Crespi and M. New. *Cultured aquatic species fact sheets*. FAO, 2009. URL https://www.fao.org/fishery/docs/DOCUMENT/aquaculture/CulturedSpecies/file/en/en_giltheadseabr.htm.

I. Cunha and M. Planas. Ingestion rates of turbot larvae (scophthalmus maximus) using different-sized live prey. In *ICES MSS Vol. 201-Mass Rearing of Juvenile Fish*, 1995. URL https://doi.org/10.17895/ices.pub.19271516.

S. R. Datta, D. J. Anderson, K. Branson, P. Perona, and A. Leifer. Computational neuroethology: a call to action. *Neuron*, 104(1):11–24, 2019. doi: 10.1016/j.neuron.2019.09.038. URL https://doi.org/10.1016/j.neuron.2019.09.038.

G. Englund and S. D. Cooper. Scale effects and extrapolation in ecological experiments. volume 33 of *Advances in Ecological Research*, pages 161–213. Academic Press, 2003. doi: https://doi.org/10.1016/S0065-2504(03)33011-9. URL https://www.sciencedirect.com/science/article/pii/S0065250403330119.

H. Fan, T. Murrell, H. Wang, K. V. Alwala, Y. Li, Y. Li, B. Xiong, N. Ravi, M. Li, H. Yang, J. Malik, R. Girshick, M. Feiszli, A. Adcock, W.-Y. Lo, and C. Feichtenhofer. PyTorchVideo: A deep learning library for video understanding. In *Proceedings of the 29th ACM International Conference on Multimedia*, 2021. https://pytorchvideo.org/.

C. Feichtenhofer, H. Fan, J. Malik, and K. He. Slowfast networks for video recognition. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6202–6211, 2019. URL https://doi.org/10.1109/10.1109/ICCV.2019.00630.

C. J. Foley, D. L. Bradley, and T. O. Höök. A review and assessment of the potential use of rna: Dna ratios to assess the condition of entrained fish larvae. *Ecological Indicators*, 60:346–357, 2016. URL https://doi.org/10.1016/j.ecolind.2015.07.00.

R. Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015. URL https://doi.org/10.1109/ICCV.2015.169.

R. Goyal, S. Ebrahimi Kahou, V. Michalski, J. Materzynska, S. Westphal, H. Kim, V. Haenel, I. Fruend, P. Yianilos, M. Mueller-Freitag, et al. The" something something" video database for learning and evaluating visual common sense. In *Proceedings of the IEEE international conference on computer vision*, pages 5842–5850, 2017. URL https://doi.ieeecomputersociety.org/10.1109/ICCV.2017.622.

J. Haladjian, Z. Hodaie, S. Nüske, and B. Brügge. Gait anomaly detection in dairy cattle. In *Proceedings of the Fourth International Conference on Animal-Computer Interaction*, pages 1–8, 2017. URL https://doi.org/10.1145/3152130.3152135.

J. A. Hanley et al. Receiver operating characteristic (roc) methodology: the state of the art. *Crit Rev Diagn Imaging*, 29(3):307–335, 1989.

R. J. Hanscom, D. L. DeSantis, J. L. Hill, T. Marbach, J. Sukumaran, A. F. Tipton, M. L. Thompson, T. E. Higham, and R. W. Clark. How to study a predator that only eats a few meals a year: high-frequency accelerometry to quantify feeding behaviours of rattlesnakes (crotalus spp.). *Animal Biotelemetry*, 11(1):1–12, 2023. URL https://doi.org/10.1186/s40317-023-00332-3.

J. Hjort. Fluctuations in the great fisheries of northern europe viewed in the light of biological research. *Rapports et Procès-Verbaux des Réunions*, 20:1–13, 1914.

R. Holzman, V. China, S. Yaniv, and M. Zilka. Hydrodynamic constraints of suction feeding in low reynolds numbers, and the critical period of larval fishes. *Integrative and comparative biology*, 55(1):48–61, 2015. doi: 10.1093/icb/icv030. URL https://doi.org/10.1093/icb/icv030.

E. D. Houde. Emerging from hjort's shadow. *Journal of Northwest Atlantic Fishery Science*, 41, 2008. doi: 10.2960/J.v41.m634. URL https://doi.org/10.2960/J.v41.m634.

E. D. Houde and R. C. Schekter. Feeding by marine fish larvae: developmental and functional responses. *Environmental Biology of Fishes*, 5(4): 315–334, 1980. doi: 10.1007/BF00005186. URL https://doi.org/10.1007/BF00005186.

D. Joska, L. Clark, N. Muramatsu, R. Jericevich, F. Nicolls, A. Mathis, M. W. Mathis, and A. Patel. Acinoset: a 3d pose estimation dataset and baseline models for cheetahs in the wild. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 13901–13908. IEEE, 2021. URL https://doi.org/10.1109/ICRA48506.2021.9561338.

H. O. Kamacı, Ş. Saka, and K. Fırat. The cleavage and embryonic phase of gilthead sea bream (sparus aurata linnaeus, 1758) eggs. *Su Ürünleri Dergisi*, 22(1):205–209, 2005. URL http://www.egejfas.org/en/pub/issue/5017/68075?publisher=ege.

J. Kay, P. Kulits, S. Stathatos, S. Deng, E. Young, S. Beery, G. Van Horn, and P. Perona. The caltech fish counting dataset: A benchmark for multiple-object tracking and counting. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VIII*, pages 290–311. Springer, 2022. URL https://doi.org/10.1007/978-3-031-20074-8_17.

A. Kiersztyn, P. Karczmarek, R. Łopucki, K. Kiersztyn, T. Nowicki, K. Perzanowski, and W. Olech. The use of information granules to detect anomalies in spatial behavior of animals. *Ecological Indicators*, 136:108583, 2022. URL https://doi.org/10.1016/j.ecolind.2022.108583.

L. Koch, I. Shainer, T. Gurevich, and R. Holzman. The expression of agrp1, a hypothalamic appetite-stimulating neuropeptide, reveals hydrodynamic-induced starvation in a larval fish. *Integrative Organismal Biology*, 1(1):oby003, 2019. URL https://doi.org/10.1093/iob/oby003.

M. Kowal, M. Siam, M. A. Islam, N. D. Bruce, R. P. Wildes, and K. G. Derpanis. A deeper dive into what deep spatiotemporal networks encode: Quantifying static vs. dynamic information. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13999–14009, 2022. URL https://doi.org/10.1109/CVPR52688.2022.01361.

R. Labuguen, J. Matsumoto, S. B. Negrete, H. Nishimaru, H. Nishijo, M. Takada, Y. Go, K.-i. Inoue, and T. Shibata. Macaquepose: a novel "in the wild" macaque monkey pose dataset for markerless motion capture. *Frontiers in behavioral neuroscience*, 14:581154, 2021. URL https://doi.org/10.3389/fnbeh.2020.581154.

Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *nature*, 521(7553):436–444, 2015. doi: 10.1038/nature14539. URL https://doi.org/10.1038/nature14539.

M. Lenning, J. Fortunato, T. Le, I. Clark, A. Sherpa, S. Yi, P. Hofsteen, G. Thamilarasu, J. Yang, X. Xu, et al. Real-time monitoring and analysis of zebrafish electrocardiogram with anomaly detection. *Sensors*, 18(1):61, 2017. URL https://doi.org/10.3390/s18010061.

B. MacKenzie, W. Leggett, R. Peters, et al. Estimating larval fish ingestion rates: Can laboratory derived values be reliably extrapolated to the wild? *Marine ecology progress series. Oldendorf*, 67(3):209–225, 1990. URL https://doi.org/10.3354/meps067209.

B. R. MacKenzie and T. Kiørboe. Larval fish feeding and turbulence: a case for the downside. *Limnology and Oceanography*, 45(1):1–10, 2000. URL https://doi.org/10.4319/lo.2000.45.1.0001.

C. H. Martin and P. C. Wainwright. Multiple fitness peaks on the adaptive landscape drive adaptive radiation in the wild. *Science*, 339(6116): 208–211, 2013. URL https://www.science.org/10.1126/science.1227710.

A. Mathis, P. Mamidanna, K. M. Cury, T. Abe, V. N. Murthy, M. W. Mathis, and M. Bethge. Deeplabcut: markerless pose estimation of user-defined body parts with deep learning. *Nature neuroscience*, 21(9):1281–1289, 2018. doi: 10.1038/s41593-018-0209-y. URL https://doi.org/10.1038/s41593-018-0209-y.

J. B. McGraw and H. Caswell. Estimation of individual fitness from life-history data. *The American Naturalist*, 147(1):47–64, 1996. URL https://doi.org/10.1086/285839.

E. M. McLean, E. A. Archie, and S. C. Alberts. Lifetime fitness in wild female baboons: trade-offs and individual heterogeneity in quality. *The American Naturalist*, 194(6):745–759, 2019. URL https://doi.org/10.1086/705810.

A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural*

Information Processing Systems 32, pages 8024–8035. Curran Associates, Inc., Vancouver, Canada, 2019. URL http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

P. Pepin. Feeding by larval fish: how taxonomy, body length, mouth size, and behaviour contribute to differences among individuals and species from a coastal ecosystem. *ICES Journal of Marine Science*, 80(1):91–106, 2023. URL https://doi.org/10.1093/icesjms/fsac215.

T. Saito and M. Rehmsmeier. The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, 10(3):e0118432, 2015. doi: 10.1371/journal.pone.0118432. URL https://doi.org/10.1371/journal.pone.0118432.

T. Saito and M. Rehmsmeier. Precrec: fast and accurate precision–recall and roc curve calculations in r. *Bioinformatics*, 33(1):145–147, 2017. doi: 10.1093/bioinformatics/btw570. URL https://doi.org/10.1093/bioinformatics/btw570.

B. L. Sargeant, J. Mann, P. Berggren, and M. Krützen. Specialization and development of beach hunting, a rare foraging behavior, by wild bottlenose dolphins (tursiops sp.). *Canadian Journal of Zoology*, 83(11):1400–1410, 2005. URL https://doi.org/10.1139/z05-136.

C. Schlechtriem, U. Focken, and K. Becker. Stable isotopes as a tool for nutrient assimilation studies in larval fish feeding on live food. *Aquatic Ecology*, 38:93–100, 2004. URL https://dx.doi.org/10.1023/b:aeco.0000020951.76155.3e.

D. Schluter. Ecology and the origin of species. *Trends in ecology & evolution*, 16(7):372–380, 2001. URL https://doi.org/10.1016/S0169-5347(01)02198-X.

E. Shamur, M. Zilka, T. Hassner, V. China, A. Liberzon, and R. Holzman. Automated detection of feeding strikes by larval fish using continuous high-speed digital video: a novel method to extract quantitative data from fast, sparse kinematic events. *Journal of Experimental Biology*, 219 (11):1608–1617, 2016. doi: 10.1242/jeb.133751. URL https://doi.org/10.1242/jeb.133751.

E. V. Sheehan, D. Bridger, S. J. Nancollas, and S. J. Pittman. Pelagicam: A novel underwater imaging system with computer vision for semi-automated monitoring of mobile marine fauna at offshore structures. *Environmental monitoring and assessment*, 192:1–13, 2020. URL https://doi.org/10.1007/s10661-019-7980-4.

R. Shields. Larviculture of marine finfish in europe. *Aquaculture*, 200(1-2):55–88, 2001. doi: 10.1016/S0044-8486(01)00694-9. URL https://doi.org/10.1016/S0044-8486(01)00694-9.

N. Sommerfeld and R. Holzman. The interaction between suction feeding performance and prey escape response determines feeding success in larval fish. *Journal of Experimental Biology*, 222(17):jeb204834, 2019. doi: 10.1242/jeb.204834. URL https://doi.org/10.1242/jeb.204834.

C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu. A survey on deep transfer learning. In V. Kůrková, Y. Manolopoulos, B. Hammer, L. Iliadis, and I. Maglogiannis, editors, *Artificial Neural Networks and Machine Learning – ICANN 2018*, pages 270–279, Cham, 2018. Springer International Publishing. ISBN 978-3-030-01424-7. URL https://doi.org/10.1007/978-3-030-01424-7_27.

Y. Tanaka, K. Satoh, H. Yamada, T. Takebe, H. Nikaido, and S. Shiozawa. Assessment of the nutritional status of field-caught larval pacific bluefin tuna by rna/dna ratio based on a starvation experiment of hatchery-reared fish. *Journal of Experimental Marine Biology and Ecology*, 354(1): 56–64, 2008. URL https://doi.org/10.1016/j.jembe.2007.10.007.

J. A. Thomson, A. Gulick, and M. R. Heithaus. Intraspecific behavioral dynamics in a green turtle chelonia mydas foraging aggregation. *Marine Ecology Progress Series*, 532:243–256, 2015. URL https://doi.org/10.3354/meps11346.

D. Tuia, B. Kellenberger, S. Beery, B. R. Costelloe, S. Zuffi, B. Risse, A. Mathis, M. W. Mathis, F. van Langevelde, T. Burghardt, et al. Perspectives in machine learning for wildlife conservation. *Nature communications*, 13(1):1–15, 2022. doi: 10.1038/s41467-022-27980-y. URL https://doi.org/10.1038/s41467-022-27980-y.

J. Van de Walle, B. Larue, G. Pigeon, and F. Pelletier. Different proxies, different stories? imperfect correlations and different determinants of fitness in bighorn sheep. *Ecology and Evolution*, 12(12):e9582, 2022. URL https://doi.org/10.1002/ece3.9582.

V. A. Viblanc, C. Saraux, A. Tamian, F. Criscuolo, D. W. Coltman, S. Raveh, J. O. Murie, and F. S. Dobson. Measuring fitness and inferring natural selection from long-term field studies: different measures lead to nuanced conclusions. *Behavioral Ecology and Sociobiology*, 76(6):79, 2022. URL https://doi.org/10.1007/s00265-022-03176-8.

B. G. Weinstein. A computer vision for animal ecology. *Journal of Animal Ecology*, 87(3):533–545, 2018. doi: 10.1111/1365-2656.12780. URL https://doi.org/10.1111/1365-2656.12780.

R. W. Wrangham and E. Van Zinnicq Bergmann Riss. Rates of predation on mammals by gombe chimpanzees, 1972–1975. *Primates*, 31:157–170, 1990. URL https://doi.org/10.1007/BF02380938.

Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick. Detectron2. https://github.com/facebookresearch/detectron2, 2019.

S. Xiao, Y. Wang, A. Perkes, B. Pfrommer, M. Schmidt, K. Daniilidis, and M. Badger. Multi-view tracking, re-id, and social network analysis of a flock of visually similar birds in an outdoor aviary. *International Journal of Computer Vision*, 131(6):1532–1549, 2023. URL https://doi.org/10.1007/s11263-023-01768-z.

S. Yaniv, D. Elad, and R. Holzman. Suction feeding across fish life stages: flow dynamics from larvae to adults and implications for prey capture. *Journal of Experimental Biology*, 217(20):3748–3757, 2014. URL https://doi.org/10.1242/jeb.104331.

M. Yúfera and M. Darias. The onset of exogenous feeding in marine fish larvae. *Aquaculture*, 268(1-4):53–63, 2007. URL https://doi.org/10.1016/j.aquaculture.2007.04.050.

M. Yúfera, F. J. Moyano, and G. Martínez-Rodríguez. The digestive function in developing fish larvae and fry. from molecular gene expression to enzymatic activity. *Emerging issues in fish larvae research*, pages 51–86, 2018. URL https://doi.org/10.1007/978-3-319-73244-2_3.

A. Zeileis, C. Kleiber, and S. Jackman. Regression models for count data in R. *Journal of Statistical Software*, 27(8), 2008. URL http://www.jstatsoft.org/v27/i08/.